Author Response

We'd like to thank the editor for catching this issue – we agree that the mole density should be that at the emission location not the sampling location. We have re-run the model to produce the appropriate output to enable this calculation and repeated our analysis. The corrected results have been updated in the text and tables, and we have replotted Figs. 2, 3, 4 and 6 with the new simulated data. The changes do not affect the conclusions so the only significant textual change is in the description of the model run itself.

Assessing London CO_2, CH_4 and CO emissions using aircraft measurements and dispersion modelling

Joseph R. Pitt[1], Grant Allen[1], Stéphane J.-B. Bauguitte[2], Martin W. Gallagher[1], James D. Lee[3], Will Drysdale[3], Bethany Nelson[3], Alistair J. Manning[4], Paul I. Palmer[5]

[1]School of Earth and Environmental Sciences, University of Manchester, Oxford Road, Manchester, M13 9PL, UK
[2]Facility for Airborne Atmospheric Measurements (FAAM), Cranfield University, Cranfield, MK43 0AL, UK
[3]Wolfson Atmospheric Chemistry Laboratories, Department of Chemistry, University of York, Heslington, York YO10 5DD, UK
[4]Met Office, FitzRoy Road, Exeter, EX1 3PB, UK
[5]School of GeoSciences, University of Edinburgh, Edinburgh, EH9 3FF, UK

Correspondence to: Joseph Pitt (joseph.pitt@manchester.ac.uk)

Abstract. We present a new modelling approach for assessing atmospheric emissions from a city, using an aircraft measurement sampling strategy similar to that employed by previous mass balance studies. Unlike conventional mass balance methods, our approach does not assume that city-scale emissions are confined to a well-defined urban area and that peri-urban emissions are negligible. We apply our new approach to a case study conducted in March 2016, investigating CO, CH_4 and CO_2 emissions from a region focussed around Greater London using aircraft sampling of the downwind plume. For each species, we simulate the flux per unit area that would be observed at the aircraft sampling locations based on emissions from the UK national inventory, transported using a Lagrangian dispersion model. To reconcile this simulation with the measured flux per unit area, assuming the transport model is not biased, we require that inventory values of CO, CH_4, and CO_2 are scaled by 1.0~~3~~0, 0.7~~1~~0, and 1.6~~1~~57, respectively. However, our result for CO_2 should not be considered a direct comparison with the inventory which only includes anthropogenic fluxes.

For comparison, we also calculate fluxes using a conventional mass balance approach and compare these to the emissions inventory aggregated over the Greater London area. Using this method we derive much higher inventory scale factors for all three gases, as a direct consequence of the failure to account for emissions outside the Greater London boundary. That substantially different conclusions are drawn using the conventional mass balance method demonstrates the danger of using this technique for cities whose emissions cannot be separated from significant surrounding sources.

1. Introduction

Over half the people in the world (54%) live in urban areas. This proportion is projected to increase to 66% by 2050 (United Nations, 2014). Consequently, cities are responsible for a large proportion of anthropogenic greenhouse gas (GHG) emissions. The 2015 UNFCCC Paris Agreement requires signatory states to not only report national GHG emissions, but also to establish and improve independent methods for verifying these reported emissions (UNFCCC, 2015). Top-down

methods that use atmospheric measurements to determine city-scale emissions can assess the accuracy of bottom-up emission inventories and provide crucial information to help improve bottom-up accounting methods.

In the UK, spatially and sectorally disaggregated emissions calculated using a bottom-up methodology are given in the National Atmospheric Emissions Inventory (NAEI; Brown et al., 2017). For Greater London, nearly all sources of anthropogenic CO_2 and CO emissions are associated with fuel combustion. For CO_2, the main sources are domestic and commercial combustion and road transport, while emissions from power stations are largely located outside the Greater London administrative boundary. Emissions of CO are comprised of a range of combustion sources, with road transport emissions constituting the largest category. For CH_4 the principal sources are waste treatment and disposal, and leakage during natural gas distribution (in contrast to the UK as a whole, where emissions associated with ruminant livestock dominate). Providing a top-down constraint on these London emissions is important, not only because London represents a large emission source in its own right, but also because it can help inform the assumptions that go into calculating bottom-up emission estimates for these sectors at a national level.

Natural emissions, which are not included in the NAEI, contribute to varying extents for the three species. Wetlands are the most significant source of natural CH_4 emissions within the UK, but wetland fluxes from London and its surrounding areas are negligible compared to anthropogenic emissions. Likewise, in urban areas CO is dominated by anthropogenic sources, although oxidation of biogenic VOCs can contribute, especially during the summer. However, biospheric fluxes do have a significant impact on measured CO_2 mole fractions downwind of the UK; the impact of these fluxes is discussed further in Sect. 3.1.2.

Aircraft mass balance techniques have previously been employed to measure trace gas fluxes from several cities (e.g. Mays et al., 2009; Turnbull et al., 2011; Cambaliza et al., 2014) including London (O'Shea et al., 2014a). Typically, horizontal transects are conducted downwind of a city at several altitudes to sample its emitted plume for various trace gas species. These vertically-stacked transects define a 2D plane of sampling downwind of the city. A background mole fraction can be determined by sampling upwind of the city or from downwind measurements outside of the plume. The mass flux of the plume through the 2D plane of sampling is then calculated from the measured mole fraction enhancements (above background) and wind speed. This approach works well for isolated cities, but for cities such as London that are surrounded by other emission sources it is difficult to measure a background that allows direct comparison between the mass balance flux and inventory emissions aggregated over a well-defined area. The impact of this issue is one of the key focus points of this study.

Another approach to flux quantification involves the use of an atmospheric transport model to represent transport of emitted species from source to measurement site. This enables simulated enhancements to be calculated at the measurement location based on a prescribed emissions map (e.g. from a bottom-up inventory). A range of inverse modelling techniques can then be employed to optimise the emissions map according to the measured mole fractions. This is frequently performed at a regional scale using ground-based measurements from long-term monitoring sites (e.g. Manning et al., 2011; Bergamaschi et al., 2015; Ganesan et al., 2015), but it has also been performed using aircraft data to provide the spatial

sampling coverage required to estimate city-scale emissions using a few hours of measurement data (e.g. Brioude et al., 2013).

The approach taken to performing the inversion must be tailored to the measurement dataset. It is important to allow the inversion sufficient freedom such that significant differences between measured and modelled mole fractions are reflected in differences between the posterior and prior emission maps. However, allowing too much freedom can result in unrealistically large redistributions of emissions in the posterior solution driven by errors in either model transport or measurements. Striking this balance is particularly difficult when using aircraft data, which has a greatly reduced temporal coverage compared to continuous ground-based measurements. Inversions using ground-based measurements are typically performed on monthly to annual timescales; the systematic biases in model transport over these timescales are greatly reduced compared with the model transport error at the time of a given aircraft flight.

In this study we have developed a new approach for assessing bottom-up inventory fluxes, using the same aircraft sampling technique typically employed by mass balance studies. This method uses the UK Met Office's Lagrangian dispersion model, NAME (Numerical Atmospheric dispersion Modelling Environment), to simulate the transport of inventory fluxes to the location of the measurements. We perform a simple inversion by comparing the average measured and simulated fluxes at the aircraft sample locations and rescaling the inventory according to their ratio. This approach has two main advantages over traditional inversion techniques. Firstly, it effectively controls the inversion behaviour by removing the freedom to spatially redistribute emissions, thus allowing a solution to be derived whose magnitude is completely independent from the magnitude of emissions in the prior. Secondly, by comparing fluxes rather than mole fractions at the measurement locations, the sensitivity of the results to biases in the modelled wind speed is reduced. On the other hand, relative to the mass balance method, we are able to account for the presence of sources outside the Greater London boundary such that they do not bias our conclusions.

We demonstrate the implementation of this method by applying it to a single flight case study downwind of London, conducted in March 2016. In addition to applying our new method, we also apply the conventional mass balance technique to the same data, and compare the top-down fluxes derived to inventory emissions aggregated over the Greater London administrative area. We discuss the differences between the results from both methods and reflect on the appropriate context in which each can be applied.

2. Case study details

2.1 Aircraft measurements and calibration

We recorded measurements on board the UK's Facility for Airborne Atmospheric Measurement (FAAM) BAe-146 atmospheric research aircraft (henceforth referred to as the FAAM aircraft). For full details of the aircraft payload see Palmer et al. (2018). Here we describe only those measurements that are relevant to this case study.

Mole fractions of CO_2 and CH_4 were measured using a Fast Greenhouse Gas Analyser (FGGA; Los Gatos Research, USA). Paul et al. (2001) describe the operating principle of the instrument, and O'Shea et al. (2013) provide details on the instrument operational practice and performance on the FAAM aircraft across several campaigns. The FGGA was calibrated hourly in flight, using two calibration gas standards traceable to the WMO-X2007 scale (Tans et al., 2009) and WMO-X2004A scale (an extension of the scale described by Dlugokencky et al., 2005) for CO_2 and CH_4, respectively. For this case study, the certified standards (369.54 ppm CO_2, 1853.8 ppb CH_4 and 456.97 CO_2, 2566.0 ppb CH_4 respectively) spanned the range of measured ambient mole fractions.

We measured a target cylinder containing intermediate mole fraction values approximately half way between these hourly calibrations to quantify any instrument non-linearity or drift. This flight formed part of the wider GAUGE (Greenhouse gAs UK and Global Emissions) campaign, during which we derived average target cylinder measurement offsets of 0.036 ppm for CO_2 and 0.07 ppb for CH_4, relative to the WMO-traceable values, with standard deviations of 0.398 ppm and 2.40 ppb, respectively, for 1 Hz sampling. Each individual target cylinder measurement consisted of a 20 s sample after allowing time for the measurements to reach equilibrium. The standard deviation of these 20 s averaged values was 0.245 ppm for CO_2 and 1.42 ppb for CH_4.

Another source of measurement uncertainty was the impact of water vapour in the sampled air on the retrieved CO_2 and CH_4 mole fractions. This was principally a consequence of mole fraction dilution (i.e. an increase in the total number of molecules per unit volume relative to dry air) and pressure broadening of the spectral absorption lines. The method used to correct for this is described by O'Shea et al. (2013). Using that technique we have derived maximal uncertainties due water vapour of 0.156 ppm for CO_2 and 1.05 ppb for CH_4. Finally, there are also uncertainties associated with the certification of the target cylinder of 0.075 ppm and 0.76 ppb, respectively. Combining all of these uncertainties with the target measurement standard deviations yields nominal total uncertainties for CO_2 and CH_4 of 0.434 ppm and 2.73 ppb at 1 Hz, and 0.300 ppm and 1.93 ppb when averaged over 20 s.

We measured CO mole fractions using vacuum ultraviolet florescence spectroscopy (AL5002, Aerolaser GmbH, Germany). The principle of this system is described by Gerbig et al. (1999), who also evaluate its performance on board an aircraft. Calibration was performed using in-flight measurements of a single gas standard and the background signal at zero CO mole fraction. Gerbig et al. (1999) derive a 1 Hz repeatability of 1.5 ppb (at 100 ppb), and an accuracy of 1.3 ppb ± 2.4% for the 1 Hz measurements.

Details of the meteorological instrumentation on board the FAAM aircraft are provided by Petersen and Renfrew (2009). In summary, we measured temperature with a Rosemount 102AL sensor, with an overall measurement uncertainty of 0.3 K at 95% confidence; we took static pressure measurements from the air data computer, based on measurements from pitot tubes around the fuselage, with an estimated absolute accuracy of 0.5 hPa; we made 3D wind measurements using the 5-hole probe system described by Brown et al. (1983), with an estimated uncertainty in horizontal wind measurements of < 0.5 m s^{-1}.

2.2 Sampling strategy

On 04 March 2016 we conducted a targeted case study flight to measure CO_2, CH_4 and CO mole fraction enhancements downwind of London, so as to assess the accuracy of the bottom-up emissions inventory. The flow over the region was consistently westerly, bringing background air from the northern Atlantic with an average travel time over the British Isles of 20 hours (as determined using NAME). As the influence of land-based sources on the recent history of this air mass (upwind of the British Isles) can be assumed to be negligible, we expect that air arriving at the British Isles had relatively homogeneous and well-mixed trace gas composition throughout the boundary layer prior to the influence of local fluxes. Take off was at 08:55 local time (=UTC), with the vertically-stacked transects downwind of London conducted between 11:16 and 13:32 local time. Results from NAME indicate a typical travel time between central London and the downwind sampling plane of ~5 hours, suggesting the majority of the sampled air passed over London between ~06:00 and ~09:00.

Figure 1 shows the flight track from an aerial perspective; between points A and B we flew repeated horizontal transects at various altitudes through a plume of enhanced mole fractions downwind of London emission sources. At the southernmost end of these transects, the constraints of UK airspace forced us to deviate from our desired course perpendicular to the prevailing wind. However, as we sampled the overwhelming majority of the London plume north of this imposed turning point, such that measurements during the deviation to point B represented background (out-of-plume) sampling, we do not expect this deviation to impact on the derived fluxes.

During an initial transect at 1550 m altitude we measured typical uniform free tropospheric background mole fractions for all three gases (CO_2, CH_4 and CO). Following this we descended to 120 m, and subsequently flew 6 transects of increasing altitude, breaking the final transect short to profile up to 1550 m within the observed plume. Figures 2a, 2b and 2c show these transects, coloured by mole fraction for CO_2, CH_4 and CO respectively, projected onto an altitude-latitude plane.

2.3 Dispersion model configuration

To determine the air history corresponding to the continuous aircraft sampling we ran the NAME dispersion model in backwards mode, releasing 100 tracer particles at each 1 Hz aircraft measurement location and tracking their motion back in time. NAME was driven by meteorological data from the UK Met Office's UKV model (Tang et al., 2013), which provides hourly data on 70 vertical levels at 1.5 km horizontal resolution over the British Isles. NAME determines particle motion based on the mean wind field (which is determined by interpolating the met data spatially and temporally to the particle location for each time-step) and a parameterisation of unresolved turbulent and mesoscale motions (for details see Jones et al., 2007, and references therein). In this study we used a NAME model time-step of 1 minute. By way of guidance, it is worth noting that although this NAME setup is more computationally intensive than is typically employed, because the release duration was less than 3 hours and the maximum particle travel time before leaving the domain was 37 hours (and less than 24 hours for the majority of particles) the run completion time remained on the order of hours rather than days using the JASMIN scientific computing facility.

To quantify the sensitivity of the sampling to surface fluxes, we used NAME to calculate an air history matrix for each minute of the flight (henceforth referred to as a release period). Each tracer particle was assigned a nominal mass and molar mass, enabling NAME to calculate the volumetric mixing ratio of tracer ~~We recorded the total time each tracer particle spent~~ within the lowest 100 m above ground level on a 1 km x 1 km horizontal grid (UK National Grid)~~, matching~~. This was chosen to match the spatial resolution of the NAEI emissions inventory. A different tracer label was used for each release period, and the model output for each release period was provided as a time-integrated volumetric mixing ratio, summed over all time-steps. The air history matrix D_{ij} was then calculated according to:

$$D_{ij}(t) = \frac{A_{ij} \times r_{ij}(t)}{N_{tot}} \quad \sout{D_{ij}(t) = \frac{T_{ij}(t)}{d \times n(t)}} \qquad (1)$$

The indices i and j represent the northing and easting components of the horizontal grid. The time-integrated mixing ratio of tracer in each grid square is given by $r_{ij}(t)$, the area of each grid cell (here 1 km^2) by A_{ij} and the total moles of tracer in a single release period by N_{tot}. ~~$T_{ij}(t)$ is the average particle residence time in each grid box for every release period, d is the height of the grid boxes (here 100 m) and $n(t)$ is the average molar density of air (mol$_{air}$ m^{-3}) at the aircraft sampling locations based on UKV data.~~

This air history matrix represents the mole fraction enhancement at the sample locations due to a unit flux in each grid box. By combining this information with the NAEI inventory emissions (F_{ij}) we can calculate a time series of simulated mole fraction enhancements (X) at the aircraft sample locations:

$$X(t) = \sum_{i,j} D_{ij}(t) \times F_{ij} \qquad (2)$$

The emissions F_{ij} are given here in units of mol$_X$ m^{-2} s^{-1}, where X represents either CO_2, CH_4 or CO. Figures 2d, 2e and 2f show the equivalent data to Figs. 2a, 2b and 2c, coloured by simulated mole fraction enhancement rather than measured mole fraction.

3. Inventory flux comparisons

In this section we present two approaches to assess the accuracy of the NAEI inventory emissions relative to the measured mole fractions during this case study. The first is a new approach, referred to hereafter as the flux-dispersion method, using the simulated mole fraction enhancements from Sect. 2.3 to derive simulated fluxes through the downwind sampling plane based on inventory emissions, thus enabling comparison with corresponding measured fluxes. The results from this method represent our best assessment of inventory fluxes for this case study.

In Sect. 3.2 we then employ a conventional mass balance method to derive top-down fluxes which are compared to an aggregated NAEI value. We discuss the outcomes of both approaches in Sect. 3.3 and explain how the conventional mass balance approach can lead to spurious conclusions in cases such as this.

It is important to note that the NAEI contains only annually averaged emissions and so does not capture the potentially large temporal variability on diurnal, weekly and seasonal timescales. Clearly this represents a likely source of difference between the top-down results derived from our single flight case study (which represent a snapshot in time) and the inventory. The most recent gridded emissions available in the NAEI at the time of writing were for the year 2015; therefore we have used these 2015 emissions to represent the 2016 values in both approaches. The UK totals (not spatially disaggregated) for 2016 have been released, allowing us to compare these to the 2015 totals. For CO there was a 9.4% reduction in total reported emissions between 2015 and 2016, while for CO_2 there was a reduction of 5.8% and for CH_4 there was an increase of 0.01%. These inter-annual changes are likely to be small in comparison with the variability on shorter timescales mentioned above.

3.1 Flux-dispersion method

3.1.1 Methodology

To make a comparison between the measured and simulated datasets described in Sect. 2 it is first necessary to calculate a background mole fraction for both, so that the mole fraction enhancement due to the London plume can be determined. To determine periods of sampling that were not significantly influenced by the London plume, and therefore can be considered to represent background mole fractions, we again utilised the air history information given by the NAME dispersion modelling. From the gridded air histories described in Sect. 2.3, we calculated the fraction of $D_{ij}(t)$ that was within the Greater London administrative region for each release period, and defined all release periods where this fraction was less than 0.05% as background periods. This Greater London fraction is shown in Fig. 3, with the background periods coloured red.

In practice there is no sharp distinction between in-plume and background sampling, so any criteria used to separate sampling into these two categories inherently involves some level of human judgement. The key consideration when defining the background for use with this method is to use a threshold that optimises the sensitivity of the results to the region of interest, in this case Greater London. This is illustrated by Fig. 4, which shows the air history ($D_{ij}(t)$) aggregated over both the background (Fig 4a) and in-plume periods (Fig 4b). This clearly illustrates that the background criteria used here avoids air histories with significant influence from the London conurbation.

The comparison between measured and simulated flux discussed later in this section is a comparison between the flux enhancement from the areas sampled in Fig. 4b relative to the flux enhancement from the air histories sampled in Fig. 4a. Clearly this comparison is not entirely selective of emissions from Greater London, with additional influence from emissions within a wider area (largely upwind and downwind of Greater London). However, given the sampling strategy employed it is not possible to isolate Greater London emissions from other upwind and downwind sources using any technique, and the 0.05% threshold employed represents the best choice to isolate sampling periods with significant Greater

London influence. The relative advantages of different sampling strategies for background determination are discussed further in the context of the mass balance method in Sect. 3.2.1.

For both the measured and simulated datasets the mole fraction enhancement due to the London plume is calculated by subtracting the background mole fraction. For each constant-altitude aircraft transect we calculated average background mole fractions to the north and south of the plume separately, for both measured and simulated datasets. We then calculated the mole fraction enhancement, $\Delta X_{London}(t)$, for both datasets using Eq. (3):

$$\Delta X_{London}(t) = X(t) - \frac{\left(\overline{X_{bgd\,N}}(z) + \overline{X_{bgd\,S}}(z)\right)}{2} \quad (3)$$

Here $X(t)$ is the mole fraction time series and $\overline{X_{bgd\,N}}(z)$ and $\overline{X_{bgd\,S}}(z)$ are the average background mole fractions to the north and south of the plume for each transect. The motivation for treating the background in this way is to capture potential latitudinal and vertical gradients in mole fraction. Vertical gradients are accounted for by calculating a separate background value for each transect, while latitudinal gradients are accounted for by the separate calculation of the north and south backgrounds. If a straight average over all background periods was used for each transect, this would be subject to potential bias in cases where there was more background sampling on one side of the plume than the other. Calculating north and south backgrounds separately as in Eq. (3) mitigates this issue. An alternative method is to interpolate the background values between the north and south edges of the plume, however due to the symmetry of the plume in this case interpolating rather than averaging had a negligible impact on the results (changing the final ratios by less than 1%). The background values used in Eq. (3) are given in Table 1.

The time series of measured and simulated mole fraction enhancements calculated using Eq. (3) are directly comparable quantities. However, the simulated mole fraction enhancements are strongly dependent on the model wind speeds (which have a direct impact on the average particle residence time in Eq. (1)). Any bias in the model wind speeds relative to the measured wind speeds consequently produces a bias in the simulated mole fraction enhancements. Figure 5 shows a comparison of modelled and measured wind speeds throughout the course of the flight. It can be seen that the model tends to overestimate wind speed within the boundary layer, particularly at lower altitudes.

In order to account for the low-biased simulated enhancements resulting from the high-biased model wind speeds, we convert both measured and simulated mole fraction enhancements into fluxes per unit area in the mean wind direction (i.e. through the downwind sampling plane) before making a comparison between them. To define a representative wind direction, we took the average of the mean UKV model wind direction and the mean measured wind direction during the sampling period. A time series of flux per unit area in this average wind direction, hereafter referred to as the flux density, was then calculated for both measured and simulated datasets using Eq. (4):

$$FD(t) = \Delta X_{London}(t) \times n_{air}(t) \times U_{\perp}(t) \quad (4)$$

The mole fraction enhancement (ΔX_{London}), molar density of air (n_{air}) and wind speed in the mean wind direction ($U_{\perp}$) were calculated independently for the measured and simulated datasets, producing flux densities (FD) in mol m^{-2} s^{-1} that are directly comparable.

3.1.2 Flux-dispersion results

Figure 6 shows a comparison between these measured and simulated flux densities as a function of latitude for each plume transect. The lowest transect from Figs. 2 and 3 (~120 m) has been excluded because no value for $\overline{X_{bgd\,N}}$ was obtained – this was the first transect conducted before the position of the plume had been fully established so its northern extent was not sampled. The top two transects have also been excluded here because they are entirely above the average boundary layer height of 759 m used by the NAME dispersion model for this simulation (which is taken directly from the UKV met data). Above this height the parameters used by NAME to describe the turbulent motion of the particles are set to fixed values resulting in poorer representation of particle dispersion. Within the boundary layer these parameters are calculated from the friction velocity and characteristic convective velocity. Therefore the flux densities calculated for transects within the model boundary layer are more accurate than those above it.

We note that the flux density enhancements for the two transects above the model boundary layer are underestimated by the simulation. A possible cause for this would be suppressed vertical mixing in the model as a result of the simplified turbulence parameterisation above this height. A full investigation into the impact of turbulence parameterisation on the vertical mixing within the NAME model would require a separate study, but we note that if the vertical mixing in the model is suppressed this could represent a potential source of bias, leading to larger simulated flux densities within the boundary layer than would in reality be produced by the inventory emissions.

A notable feature of the transects shown in Fig. 6 is that the centre of the simulated plume is consistently further north than the centre of the measured plume. This could suggest the spatial distribution of emissions within the inventory is incorrectly weighted towards sources in the north of London. Alternatively, any inaccuracy in the model wind field could lead to the simulated plume being advected to a more northerly position on the sampling plane than the measured plume. The fact that all three species exhibit the same northerly offset of the simulated plume points to the latter explanation, as each species has a different source mix, making it unlikely that they would all exhibit the same spatial bias.

In itself, the mismatched position of the measured and simulated plumes does not bias the results. This is one of the key advantages of comparing average flux densities for each transect, rather than using differences between the measured and simulated time series to optimise an emission map (e.g. using a cost function). However, if the plume position mismatch reflects an error in the transport model this does have the potential to impact the results, as it suggests the air histories for in-plume and background periods simulated by NAME may differ slightly from the actual air histories of the measurements. This is one possible reason why the simulated background CO and CH_4 mole fractions were higher to the north of the plume than to the south, when this gradient was not observed in the measured background (see Table 1). Alternatively, the higher simulated background to the north of the plume could be counteracted in the measured dataset if this air had a lower initial mole fraction before reaching the British Isles (as the simulated dataset implicitly assumes a uniform background for air entering the domain). The treatment of the background here can mitigate either issue as long as they result in linear changes in mole fraction with latitude.

The high bias of the simulated wind speeds relative to the measurements (shown in Fig. 5) is a further indication of transport model error. While we have accounted for the most obvious impact of this issue by comparing flux density rather than mole fraction, it could also result in simulated air histories which underestimate the cross-wind spread in the sample footprint. This could result in the simulation overestimating the sensitivity of in-plume sampling to emissions from Greater London, producing a low bias in the inventory scale factors. A robust quantification of the uncertainty associated with model wind field inaccuracy (incorporating both effects discussed above) would require an ensemble of NAME runs to be performed, driven by met data with perturbed wind fields. Such quantification is beyond the scope of this study, but we note that this is a potentially significant source of uncertainty in the context of the uncertainty ranges calculated below.

Having calculated time series of flux density for the measured and simulated datasets, we then calculated average flux densities for each transect altitude. We also calculated flux densities as an overall average using data from all three transects. These values are given in Table 2 for CO_2, CH_4 and CO, along with the ratios between measured and simulated flux densities. These ratios represent the factors by which the NAEI inventory needs to be multiplied in order to reproduce the measured flux densities (assuming there is no bias induced by the NAME transport). Therefore for this case study we conclude that NAEI emissions require scaling by the overall values of 1.030 for CO, 0.710 for CH_4 and 1.6157 for CO_2.

While there are small uncertainties associated with the measured mole fractions (as discussed in Sect. 2.1), the uncertainty in these overall inventory scale factors is expected to be dominated by NAME transport uncertainty. As discussed above, quantification of the uncertainties associated with the dispersion modelling would require a more involved modelling study using an ensemble of NAME runs. Here we take the range of scale factors across the different transects of 0.920-1.163 for CO, 0.661-0.79 for CH_4 and 1.410-1.8577 for CO_2 to give an indication of the total scale factor uncertainty for each species, while noting that there may be additional sources of model transport bias that are not captured by this range.

Our results suggest that to obtain simulated enhancements consistent with our measurements, the NAEI would require downscaling for CH_4, upscaling for CO_2, with ~~no~~ only a small rescaling required for CO. For CH_4 this is qualitatively consistent with past studies suggesting national NAEI CH_4 emission totals have been too high in previous years (e.g. Manning et al., 2011; Ganesan et al., 2015) but it differs from the most recent top-down verification report (O'Doherty et al., 2017), which finds good agreement between the NAEI CH_4 emission totals and continuous ground-based measurements in recent years. There are several potential reasons why the results in this study might differ from those in the verification report; these are discussed below in the context of our result for CH_4 but is should be noted that these sources of discrepancy could apply to all three species.

Temporal variability in emissions (not included in the NAEI) is an obvious source of difference between our results for CH_4 and those in the verification report. Helfter et al. (2016) used an eddy-covariance technique to determine the diurnal variability of London CH_4 emissions, and found that emissions increased by a factor of 1.9 (maximum-to-minimum ratio) between the early and late morning. However, in addition to true temporal variability (e.g. rush hour emissions), such techniques are susceptible to the complex nature of urban boundary layer evolution during these transition periods (Halios

and Barlow, 2018), which can result in overestimation of diurnal flux variability. Nevertheless, the fact that our study represents a snapshot in time is a key difference relative to the annual timescales of both the NAEI and the verification report.

The fact that our study focussed on London and its surrounding areas, while the verification report presents national scale results, represents another key difference between them. It is possible that, although NAEI emission totals agree with long term observations, the spatial distribution of these emissions is not well represented in the inventory, such that the proportion of emissions ascribed to urban areas is too large. The impact of temporal variability makes it impossible to draw such a conclusion from a single case study, however repeated flights at different times of day, week and year would enable this hypothesis to be tested.

For CO_2 we find that the NAEI would require upscaling in order to be consistent with observations. However, direct comparison with the NAEI is not appropriate for CO_2 because biospheric fluxes, which are not included in the NAEI, represent a significant influence on the measured mole fractions. These biospheric fluxes include uptake due to photosynthesis (gross primary production; GPP), emission from autotrophic respiration and emission from heterotrophic respiration. Net primary production (NPP) is calculated as the difference between photosynthetic uptake and autotrophic respiration. Hardiman et al. (2017) investigated biospheric CO_2 fluxes in Massachusetts and found higher NPP values outside the Boston conurbation. Combined with higher heterotrophic respiration in more populated areas (including human respiration), these higher rural NPP values result in a positive net biospheric flux from urban areas relative to surrounding rural areas. In our study this translates to a positive net biospheric flux within the footprint of the in-plume measurements relative to the footprint of the background measurements. As we have not accounted for this net biospheric flux in our simulated flux densities we expect them to underestimate the measured values, even if the NAEI emissions are entirely accurate.

Prior quantification of the biospheric impact on the derived scale factor would require the use of an ecosystem model, and is beyond the scope of this study. However, some inferences can be made by considering the different scale factors derived for CO and CO_2, as these species share many of the same combustion sources. Previous studies (e.g. O'Doherty et al., 2013) have used CO measurements as a proxy for anthropogenic CO_2, relying on the assumption that the inventory ratio for CO:CO_2 emissions is correct. In this case, that would imply that the difference in net biospheric flux between the in-plume and background sampling amounted to over half the corresponding difference in anthropogenic flux (comparing the scale factors of 1.030 for CO and 1.6157 for CO_2). However, while this comparison can be considered indicative of the potential order of magnitude for net biospheric flux, uncertainty in the inventory CO:CO_2 emission ratio limits our ability to use this method to obtain quantitative information on biospheric fluxes (see Turnbull et al., 2006, for further discussion on the use of CO:CO_2 ratios for this purpose).

3.2 Conventional mass balance method

3.2.1 Methodology

Detailed descriptions of the mass balance technique in the context of measuring urban GHG emissions are provided by many sources. In general, in the context of bulk area flux measurement, these sources can be categorized into two basic approaches; either the emissions are assumed to be well mixed up to a given height at which they are capped by a temperature inversion (Turnbull et al., 2011; Karion et al., 2013; Smith et al., 2015), or the vertically varying shape of the plume is derived by interpolation between transects flown at multiple altitudes (Mays et al., 2009; Cambaliza et al., 2014; O'Shea et al., 2014a), often using a kriging approach. Figures 2a, 2b and 2c clearly show that the assumptions of the first of these approaches (i.e. well mixed plumes up to a capping height) are not met in this case. We therefore adopt the latter of these approaches, and use kriging to represent the full structure of the plume. This approach necessarily assumes temporal invariance of the plume over the period of sampling: in this case ~2.5 hours.

Following the work of (Mays et al., 2009; Cambaliza et al., 2014; O'Shea et al., 2014a) we derive fluxes using Eq. (5):

$$F = \int_0^{z_{max}} \int_A^B \left(X_{ij} - X_0\right) n_{air}(z)\, U_{\perp ij}\, \mathrm{dx}\, \mathrm{dz} \quad (5)$$

Here F (mol s^{-1}) is the bulk flux for the emission source, X_{ij} is the kriged mole fraction for a given species, X_0 is the background mole fraction, $n_{air}(z)$ is the molar air density (here derived as a linear function of altitude based on measured values) and $U_{\perp ij}$ is the kriged wind speed perpendicular to the vertical sample plane across which the integral is taken.

Kriging is an interpolation method based on a stochastic Gaussian model, and is described in detail by Kitanidis (1997). It converts samples with sparse spatial coverage into a 2D grid of estimated values, with an associated grid of standard errors for these values. Here we use a modified version of the EasyKrig software (© Dezhang Chu and Woods Hole Ocean Institution) to perform the kriging; again more detail regarding the application of this software with regards to aircraft mass balance flux calculations is given by Mays et al. (2009). More detail regarding the kriging parameters used is included in the supplementary material.

The results from the kriging were output on a 20 x 29 cell grid, with a vertical resolution of 50 m and a horizontal resolution of 5 km respectively, as shown in Fig. 7. As the lowest transect was conducted at ~120 m altitude, the structure of the plume below this level was not constrained well by our sampling. Therefore the mole fractions for the lowest 100 m above ground level were taken to be the same as the kriged output for the layer at 100 – 150 m.

The background mole fraction X_0 should be chosen to best represent the mole fraction that would be measured downwind of Greater London if there were no emissions within Greater London. We determined this background for each trace gas by taking the average mole fraction over all cells within 15 km of the north or south boundary of the sample plane (i.e. the three columns at each edge of the plane in Fig. 7). This approach follows Mays et al. (2009) in determining the background from measurements in the downwind plane outside of the influence of the plume, and contrasts with the

approach used by O'Shea et al. (2014a), who instead used measurements upwind of London to determine the background. The impact of the background calculation on the results obtained using the mass balance method is discussed further in Sect. 3.3.

The background mole fractions used were 147.3 ppb for CO, 1941.6 ppb for CH_4 and 409.1 ppm for CO_2, which (despite the difference in background definition) are similar to the values used in the flux-dispersion method (see Table 1). Although we have used these average background values in our main analysis, we have also calculated fluxes using interpolated background values (as recommended by Heimburger et al., 2017) to test the sensitivity of the results to this choice of approach.

3.2.2 Mass balance results

The fluxes calculated using Eq. (5) are given in Table 3, along with 1σ uncertainties derived by combining the kriging standard errors with the uncertainty in background mole fraction, taken to be the standard deviation for all background cells used. Also given are the aggregated NAEI emissions for the Greater London administrative area for each species. We have derived inventory scale factors, in principle analogous to those in Sect. 3.1.2, by taking the ratio of these aggregated NAEI emissions to the flux calculated using Eq. (5) for each species. Using the conventional mass balance method we calculate that the NAEI requires rescaling by factors of 2.27 for CO, 1.22 for CH_4 and 3.08 for CO_2. The differences between these values and those derived in Sect. 3.1.2 are discussed in Sect. 3.3 below. Using interpolated (as opposed to average) background mole fractions slightly increases the calculated fluxes, but in all cases the difference is less than 7%. The NAEI scale factors derived using an interpolated background are calculated to be 2.33 for CO, 1.31 for CH_4 and 3.19 for CO_2.

3.3 Comparing the flux-dispersion and mass balance methods

In Sects. 3.1 and 3.2 two different methods were applied to the same dataset to derive scale factors for the NAEI inventory such that it agrees with aircraft observations. However, the scale factors derived using the flux-dispersion method are significantly lower than those derived using the conventional mass balance method. This is because one of the key elements of the mass balance method, the assumption that a city acts as an isolated emission source surrounded by areas with negligible emissions, is clearly violated in this case. In order to deal with these extraneous emissions one either needs to account for them in the background mole fraction (such that all downwind enhancements are solely a product of Greater London emissions) or include them in the aggregated inventory emission total against which the top-down flux is compared. Here we consider the issues associated with both approaches.

From Fig. 4 it is evident that measurements within the London plume are strongly influenced by sources upwind, downwind, to the north and to the south of Greater London. The mole fraction that would be observed at a given location in the presence of these sources, but the absence of emissions within the Greater London boundary, is clearly not a measureable quantity. Our calculated background, which we derive using measurements either side of the plume, is subject to greater influence from emission sources to the north and to the south of Greater London relative to the in-plume measurements,

while it fails to adequately capture emissions upwind and downwind of Greater London. There is no prior reason to assume that these two effects cancel each other out.

An alternative approach to background calculation utilises measurements upwind of the city. O'Shea et al. (2014a) use this method to calculate fluxes for Greater London, making the following implicit assumptions: 1) emissions upwind of the background measurements are well-mixed throughout the boundary layer, 2) the air history of the upwind sampling does not differ significantly from the air history of the downwind sampling, 3) entrainment of air into the boundary layer from above does not significantly impact the downwind mole fractions relative to the upwind measurements. All three of these assumptions appear dubious for the case study presented here; in particular it seems likely that there was significant entrainment of air into the boundary layer as it increased in depth throughout the morning. We also have insufficient sampling upwind of London to determine the extent to which the boundary layer can be considered well-mixed. These factors motivated our decision to use a downwind background.

Even in cases where the above assumptions are satisfied, using an upwind background does not solve the fundamental issue of extraneous emission sources. All emissions between the upwind and downwind transects, including those outside Greater London, contribute to the measured downwind enhancements to some extent. Therefore it is not possible to isolate the mole fraction enhancement due solely to Greater London emissions using either background calculation method. The influence of these surrounding emission sources could explain the large inventory upscaling factors derived by O'Shea et al. (2014a) when comparing their calculated mass balance fluxes to the NAEI totals for Greater London.

Given that it is not possible (even in principle) to isolate enhancements due to Greater London emissions, it is logical to consider over what area emissions can be considered to contribute to the calculated mass balance flux (for a given choice of background). Dispersion model air histories are frequently used to define the flux footprint when using Lagrangian mass balance techniques (e.g. O'Shea et al., 2014b) and Integrative Mass Boundary Layer techniques (e.g. Font et al., 2015). These techniques balance the change in species concentration within a column of air against the fluxes through the top, bottom and sides of the column.

An analogous approach to footprint calculation here would be to attribute the derived mass balance flux to the area given by the NAME air history for in-plume sampling. However, such an approach would be invalid because emissions from within this area also contribute to background sampling. This is evident from the overlap between the aggregate air histories for in-plume sampling (Fig. 4b) and background sampling (Fig. 4a); assuming all of the emissions from the area covered by the in-plume air history contributed directly to an enhancement above the background would yield a huge aggregate bottom-up flux that would not be representative of the calculated mass balance flux. Fundamentally, because emissions from many source areas contribute to some extent to both the in-plume and background measurements, making it unclear whether or not to include these in the aggregated inventory total, any choice of inventory aggregation area is inherently arbitrary.

In summary, it is not possible to determine a background such that all calculated enhancements result purely from Greater London emissions. Neither, for a given choice of background, can we unambiguously determine what area inventory

emissions should be aggregated over. This demonstrates the difficulty in employing this type of mass balance technique to estimate emissions from a non isolated source. Instead, the flux-dispersion method provides a good alternative in these cases because it explicitly accounts for the relative influence of all emissions on the in-plume and background sampling.

5. Conclusions

Aircraft mass balance techniques are an effective way of determining emissions from isolated sources, but they require surrounding areas to be negligible emission sources in order to yield robust results. This is a well-known assumption associated with these methods. However in the absence of alternative techniques using the same sample dataset against which the mass balance results can be compared, one is forced either to simply state this assumption as a caveat or to abandon the effort entirely.

In this study we have developed an alternative technique using a Lagrangian dispersion model to quantify the transport of inventory emissions to the aircraft sample locations, so that a direct comparison of flux per unit area can be made at the measurement locations. In contrast to the conventional mass balance technique, this method does not require cities to be isolated from surrounding emission sources, rendering it more appropriate in many cases. We have demonstrated this new technique by applying it to a single-flight case study measuring London emissions, which yielded inventory scale factors of 1.03~~0~~ (0.92~~0~~-1.16~~3~~) for CO, 0.71~~0~~ (0.66~~1~~-0.79) for CH_4 and 1.615~~7~~ (1.41~~0~~-1.85~~77~~) for CO_2. These values represent the factors by which the inventory emissions need to be multiplied to agree with the aircraft measurements, although the absence of biospheric fluxes in the inventory means direct comparison with the CO_2 measurements is not appropriate. Using a mass balance approach we derived significantly higher values (2.27, 1.22 and 3.08 respectively), which we conclude are biased as a consequence of significant sources outside the Greater London administrative region, which are neither adequately captured by the background mole fraction calculation or easy to account for in the choice of inventory aggregation area. The magnitude of this bias demonstrates how employing a mass balance method for a non-isolated source can lead to highly misleading conclusions regarding the accuracy of the emissions inventory under study.

It is important to emphasise that the inventory scale factors derived here represent the results from a single case study and therefore are not necessarily representative of the annual timescale of the NAEI emissions. In order to better validate the inventory on this timescale repeated flights following a similar sampling strategy are required. The limited spatial selectivity of the flux-dispersion technique represents another caveat on the results from a single flight, as the derived flux ratios are not only sensitive to emissions from the London conurbation but also to emissions from a fairly wide area surrounding it. Repeated flights should therefore be designed to incorporate sampling under different prevailing wind directions, so that the systematic impact of extraneous sources on the overall results is minimised. Using the flux-dispersion method developed here in combination with representative aircraft sampling on an annual timescale could provide a robust assessment of inventory fluxes at the city scale in the case of non-isolated sources for which the mass balance technique is not appropriate.

Author contributions

Conceptualization, J. P., G. A., J. L.; Data curation; J. P., S. B.; Formal analysis; J. P.; Funding acquisition; G. A., M. G., J. L., A. M., P. P.; Investigation; J. P., G. A., S. B., J. L., W. D., B. N.; Methodology; J. P., G. A.; Software, A. M.; Supervision, G. A., M. G., J. L., P. P.; Writing – original draft, J. P.; Writing – review & editing, J. P. G. A., P. P.

Acknowledgements

The authors would like to thank everyone at FAAM, Airtask and Avalon Aero who helped with the collection and processing of the aircraft data used here. We acknowledge use of the NAME atmospheric dispersion model and associated NWP meteorological data sets made available to us by the Met Office. We acknowledge the significant storage resources and analysis facilities made available to us on JASMIN by STFC CEDA along with the corresponding support teams. J. R. Pitt received a NERC CASE studentship in partnership with FAAM, grant number NE/L501/591/1, supervised by G. Allen. This work was supported by the GAUGE (Greenhouse gAs UK and Global Emissions) NERC project, grant numbers NE/K002449/1 and NE/K00221X/1.

References

Bergamaschi, P., Corazza, M., Karstens, U., Athanassiadou, M., Thompson, R. L., Pison, I., Manning, A. J., Bousquet, P., Segers, A., Vermeulen, A. T., Janssens-Maenhout, G., Schmidt, M., Ramonet, M., Meinhardt, F., Aalto, T., Haszpra, L., Moncrieff, J., Popa, M. E., Lowry, D., Steinbacher, M., Jordan, A., O'Doherty, S., Piacentino, S. and Dlugokencky, E.: Top-down estimates of European CH_4 and N_2O emissions based on four different inverse models, Atmos. Chem. Phys., 15, 715–736, doi:10.5194/acp-15-715-2015, 2015.

Brioude, J., Angevine, W. M., Ahmadov, R., Kim, S.-W., Evan, S., McKeen, S. A., Hsie, E.-Y., Frost, G. J., Neuman, J. A., Pollack, I. B., Peischl, J., Ryerson, T. B., Holloway, J., Brown, S. S., Nowak, J. B., Roberts, J. M., Wofsy, S. C., Santoni, G. W., Oda, T. and Trainer, M.: Top-down estimate of surface flux in the Los Angeles Basin using a mesoscale inverse modeling technique: assessing anthropogenic emissions of CO, NOx and CO2 and their impacts, Atmos. Chem. Phys., 13, 3661–3677, doi:10.5194/acp-13-3661-2013, 2013.

Brown, E. N., Friehe, C. A. and Lenschow, D. H.: The Use of Pressure Fluctuations on the Nose of an Aircraft for Measuring Air Motion, J. Clim. Appl. Meteorol., 22, 171–180, doi:10.1175/1520-0450(1983)022<0171:TUOPFO>2.0.CO;2, 1983.

Brown, P., Broomfield, M., Cardenas, L., Choudrie, S., Kilroy, E., Jones, L., Passant, N., Thomson, A., Wakeling, D., Buys, G., Forden, S., Gilhespy, S., Glendining, M., Gluckman, R., Henshall, P., Hobson, M., MacCarthy, J., Malcolm, H., Manning, A., Matthews, R., Milne, A., Misselbrook, T., Moxley, J., Murrells, T., Salisbury, E., Sussams, J., Thistlethwaite, G., Walker, C. and Webb, N.: UK Greenhouse Gas Inventory, 1990 to 2015, Ricardo Energy & Environment, Harwell, UK,

2017.

Cambaliza, M. O. L., Shepson, P. B., Caulton, D. R., Stirm, B., Samarov, D., Gurney, K. R., Turnbull, J., Davis, K. J., Possolo, A., Karion, A., Sweeney, C., Moser, B., Hendricks, A., Lauvaux, T., Mays, K., Whetstone, J., Huang, J., Razlivanov, I., Miles, N. L. and Richardson, S. J.: Assessment of uncertainties of an aircraft-based mass balance approach for quantifying urban greenhouse gas emissions, Atmos. Chem. Phys., 14, 9029–9050, doi:10.5194/acp-14-9029-2014, 2014.

Dlugokencky, E. J., Myers, R. C., Lang, P. M., Masarie, K. A., Crotwell, A. M., Thoning, K. W., Hall, B. D., Elkins, J. W. and Steele, L. P.: Conversion of NOAA atmospheric dry air CH4 mole fractions to a gravimetrically prepared standard scale, J. Geophys. Res., 110, D18306, doi:10.1029/2005JD006035, 2005.

Font, A., Grimmond, C. S. B., Kotthaus, S., Morguí, J.-A., Stockdale, C., O'Connor, E., Priestman, M. and Barratt, B.: Daytime CO2 urban surface fluxes from airborne measurements, eddy-covariance observations and emissions inventory in Greater London, Environ. Pollut., 196, 98–106, doi:10.1016/j.envpol.2014.10.001, 2015.

Ganesan, A. L., Manning, A. J., Grant, A., Young, D., Oram, D. . E., Sturges, W. T., Moncrieff, J. B. and O'Doherty, S.: Quantifying methane and nitrous oxide emissions from the UK and Ireland using a national-scale monitoring network, Atmos. Chem. Phys., 15, 6393–6406, doi:10.5194/acp-15-6393-2015, 2015.

Gerbig, C., Schmitgen, S., Kley, D., Volz-Thomas, A., Dewey, K. and Haaks, D.: An improved fast-response vacuum-UV resonance fluorescence CO instrument, J. Geophys. Res.-Atmos., 104, 1699–1704, doi:10.1029/1998JD100031, 1999.

Halios, C. H. and Barlow, J. F.: Observations of the Morning Development of the Urban Boundary Layer Over London, UK, Taken During the ACTUAL Project, Bound.-Lay. Meteorol., 166, 395–422, doi:10.1007/s10546-017-0300-z, 2018.

Hardiman, B. S., Wang, J. A., Hutyra, L. R., Gately, C. K., Getson, J. M. and Friedl, M. A.: Accounting for urban biogenic fluxes in regional carbon budgets, Sci. Total Environ., 592, 366–372, doi:10.1016/j.scitotenv.2017.03.028, 2017.

Heimburger, A. M. F., Harvey, R. M., Shepson, P. B., Stirm, B. H., Gore, C., Turnbull, J., Cambaliza, M. O. L., Salmon, O. E., Kerlo, A.-E. M., Lavoie, T. N., Davis, K. J., Lauvaux, T., Karion, A., Sweeney, C., Brewer, W. A., Hardesty, R. M. and Gurney, K. R.: Assessing the optimized precision of the aircraft mass balance method for measurement of urban greenhouse gas emission rates through averaging, Elem Sci Anth, 5, 26, doi:10.1525/elementa.134, 2017.

Helfter, C., Tremper, A. H., Halios, C. H., Kotthaus, S., Bjorkegren, A., Grimmond, C. S. B., Barlow, J. F. and Nemitz, E.: Spatial and temporal variability of urban fluxes of methane, carbon monoxide and carbon dioxide above London, UK, Atmos. Chem. Phys., 16, 10543–10557, doi:10.5194/acp-16-10543-2016, 2016.

Jones, A., Thomson, D., Matthew Hort and Devenish, B.: The U.K. Met Office's Next-Generation Atmospheric Dispersion Model, NAME III, in: Air Pollution Modeling and Its Application XVII, edited by C. Borrego and A.-L. Norman, 580–589, Springer US, New York, NY, USA., 2007.

Karion, A., Sweeney, C., Pétron, G., Frost, G., Hardesty, R. M., Kofler, J., Miller, B. R., Newberger, T., Wolter, S., Banta, R., Brewer, A., Dlugokencky, E., Lang, P., Montzka, S. A., Schnell, R., Tans, P., Trainer, M., Zamora, R. and Conley, S.: Methane emissions estimate from airborne measurements over a western United States natural gas field, Geophys. Res. Lett., 40, 4393–4397, doi:10.1002/grl.50811, 2013.

Kitanidis, P. K.: Introduction to Geostatistics: Applications in Hydrogeology, Cambridge University Press, Cambridge, U.K., 1997.

Manning, A. J., O'Doherty, S., Jones, A. R., Simmonds, P. G. and Derwent, R. G.: Estimating UK methane and nitrous oxide emissions from 1990 to 2007 using an inversion modeling approach, J. Geophys. Res., 116, D02305, doi:10.1029/2010JD014763, 2011.

Mays, K. L., Shepson, P. B., Stirm, B. H., Karion, A., Sweeney, C. and Gurney, K. R.: Aircraft-based measurements of the carbon footprint of Indianapolis., Environ. Sci. Technol., 43, 7816–7823, doi:10.1021/es901326b, 2009.

O'Doherty, S., Grant, A., Manning, A. J., Derwent, R. G., Simmonds, P. and Young, D.: Long-term atmospheric measurement and interpretation (of radiatively active trace gases), Report to DECC (Department of Energy and Climate Change), Contract number: GA0201, 2013.

O'Doherty, S., Stanley, K., Rigby, M., Stavert, A., Manning, A., Redington, A., Simmonds, P., Young, D. and Sturges, B.: Long-term atmospheric measurement and interpretation (of radiatively active trace gases), Report to BEIS (Department for Business, Energy and Industrial Strategy), Contract number: GA0201, 2017.

O'Shea, S. J., Bauguitte, S. J.-B., Gallagher, M. W., Lowry, D. and Percival, C. J.: Development of a cavity-enhanced absorption spectrometer for airborne measurements of CH4 and CO2, Atmos. Meas. Tech., 6, 1095–1109, doi:10.5194/amt-6-1095-2013, 2013.

O'Shea, S. J., Allen, G., Fleming, Z. L., Bauguitte, S. J.-B., Percival, C. J., Gallagher, M. W., Lee, J., Helfter, C. and Nemitz, E.: Area fluxes of carbon dioxide, methane, and carbon monoxide derived from airborne measurements around Greater London: A case study during summer 2012, J. Geophys. Res.-Atmos., 119, 4940–4952, doi:10.1002/2013JD021269, 2014a.

O'Shea, S. J., Allen, G., Gallagher, M. W., Bower, K., Illingworth, S. M., Muller, J. B. A., Jones, B. T., Percival, C. J., Bauguitte, S. J.-B., Cain, M., Warwick, N., Quiquet, A., Skiba, U., Drewer, J., Dinsmore, K., Nisbet, E. G., Lowry, D., Fisher, R. E., France, J. L., Aurela, M., Lohila, A., Hayman, G., George, C., Clark, D. B., Manning, A. J., Friend, A. D. and Pyle, J.: Methane and carbon dioxide fluxes and their regional scalability for the European Arctic wetlands during the MAMM project in summer 2012, Atmos. Chem. Phys., 14, 13159–13174, doi:10.5194/acp-14-13159-2014, 2014b.

Palmer, P. I., O'Doherty, S., Allen, G., Bower, K., Bösch, H., Chipperfield, M. P., Connors, S., Dhomse, S., Feng, L., Finch, D. P., Gallagher, M. W., Gloor, E., Gonzi, S., Harris, N. R. P., Helfter, C., Humpage, N., Kerridge, B., Knappett, D., Jones, R. L., Le Breton, M., Lunt, M. F., Manning, A. J., Matthiesen, S., Muller, J. B. A., Mullinger, N., Nemitz, E., O'Shea, S., Parker, R. J., Percival, C. J., Pitt, J., Riddick, S. N., Rigby, M., Sembhi, H., Siddans, R., Skelton, R. L., Smith, P., Sonderfeld, H., Stanley, K., Stavert, A. R., Wenger, A., White, E., Wilson, C. and Young, D.: A measurement-based verification framework for UK greenhouse gas emissions: an overview of the Greenhouse gAs Uk and Global Emissions (GAUGE) project, Atmos. Chem. Phys., 18, 11753–11777, doi:10.5194/acp-18-11753-2018, 2018.

Paul, J. B., Lapson, L. and Anderson, J. G.: Ultrasensitive absorption spectroscopy with a high-finesse optical cavity and off-axis alignment., Appl. Optics, 40, 4904–4910, 2001.

Petersen, G. N. and Renfrew, I. A.: Aircraft-based observations of air-sea fluxes over Denmark Strait and the Irminger Sea

during high wind speed conditions, Q. J. Roy. Meteor. Soc., 135, 2030–2045, doi:10.1002/qj.355, 2009.

Smith, M. L., Kort, E. A., Karion, A., Sweeney, C., Herndon, S. C. and Yacovitch, T. I.: Airborne Ethane Observations in the Barnett Shale: Quantification of Ethane Flux and Attribution of Methane Emissions, Environ. Sci. Technol., 49, 8158–8166, doi:10.1021/acs.est.5b00219, 2015.

Tang, Y., Lean, H. W. and Bornemann, J.: The benefits of the Met Office variable resolution NWP model for forecasting convection, Meteorol. Appl., 20, 417–426, doi:10.1002/met.1300, 2013.

Tans, P., Zhao, C. and Kitzis, D.: The WMO Mole Fraction Scales for CO2 and other greenhouse gases, and uncertainty of the atmospheric measurements, in: 15th WMO/IAEA Meeting of Experts on Carbon Dioxide, Other Greenhouse Gases, and Related Tracer Measurement Techniques, Jena, Germany., 2009.

Turnbull, J. C., Miller, J. B., Lehman, S. J., Tans, P. P., Sparks, R. J. and Southon, J.: Comparison of 14CO2 , CO, and SF6 as tracers for recently added fossil fuel CO2 in the atmosphere and implications for biological CO2 exchange, Geophys. Res. Lett., 33, L01817, doi:10.1029/2005GL024213, 2006.

Turnbull, J. C., Karion, A., Fischer, M. L., Faloona, I., Guilderson, T., Lehman, S. J., Miller, B. R., Miller, J. B., Montzka, S., Sherwood, T., Saripalli, S., Sweeney, C. and Tans, P. P.: Assessment of fossil fuel carbon dioxide and other anthropogenic trace gas emissions from airborne measurements over Sacramento, California in spring 2009, Atmos. Chem. Phys., 11, 705–721, doi:10.5194/acp-11-705-2011, 2011.

UNFCCC: Adoption of the Paris Agreement, in: Conference of the Parties: Twenty-first session, Paris, France., 2015.

United Nations: World Urbanization Prospects: The 2014 Revision, Highlights (ST/ESA/SER.A/352), United Nations, Department of Economic and Social Affairs, Population Division, 2014.

		CO (ppb)		CH_4 (ppb)		CO_2 (ppm)	
Dataset	Alt (m)	South	North	South	North	South	North
Measured	287	148.4	146.67.4	1943.1	1940.71.8	409.6	409.13
	460	149.2	145.3	1942.5	1938.4	409.3	408.9
	575	150.2	149.9	1943.6	1943.3	409.2	409.3
Simulated	287	3.7	4.53.9	6.8	10.21.7	0.5	0.56
	460	3.67	3.68	7.06.9	10.07	0.5	0.5
	575	3.47	4.03	6.59	11.42.7	0.45	0.5

Table 1: Background mole fractions for each species to the north and the south of the London plume, calculated using the flux-dispersion method.

	CO			CH_4			CO_2		
	Mean flux density ($\mu mol\ m^{-2}\ s^{-1}$)			Mean flux density ($\mu mol\ m^{-2}\ s^{-1}$)			Mean flux density ($\mu mol\ m^{-2}\ s^{-1}$)		
Alt (m)	Measured	Simulated	Ratio	Measured	Simulated	Ratio	Measured	Simulated	Ratio
287	1.7980	1.7480	1.030	2.703	3.416	0.79	52641	33048	1.6056
460	1.65	1.7984	0.920	2.12	3.2250	0.661	468	3314	1.410
575	1.94	1.6771	1.163	1.96	2.9185	0.679	556	30014	1.8577
Overall	1.79	1.739	**1.030**	2.28	3.1928	**0.710**	51621	32132	**1.6157**

Table 2: Mean flux densities calculated using the flux-dispersion method, given for each transect and taken over all three transects. The ratios between measured and simulated flux densities are all given.

	CO		CH_4		CO_2	
	Mean	1σ	Mean	1σ	Mean	1σ
Flux (kmol s^{-1})	0.178	0.006	0.182	0.009	44.7	1.2
NAEI emissions (kmol s^{-1})	0.079	N/A	0.149	N/A	14.5	N/A
Ratio	**2.27**	0.07	**1.22**	0.06	**3.08**	0.08

Table 3: Bulk fluxes calculated using a conventional mass balance technique and corresponding NAEI emissions, aggregated over the Greater London administrative region. The ratio of mass balance flux to NAEI emission is also given.

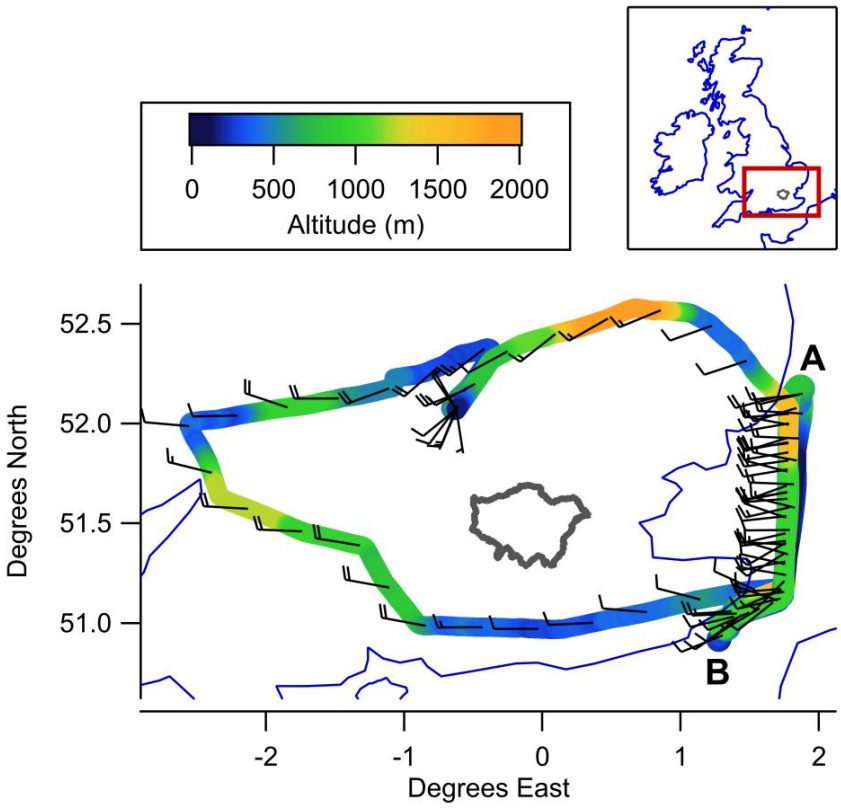

Figure 1: Aircraft flight track on 04 March 2016, coloured by altitude. Wind barbs are used to represent wind speed and direction, averaged over 5 minutes, using the convention where each full wind barb represents a wind speed of 10 knots. The border of the Greater London administrative region is shown in grey for reference.

5

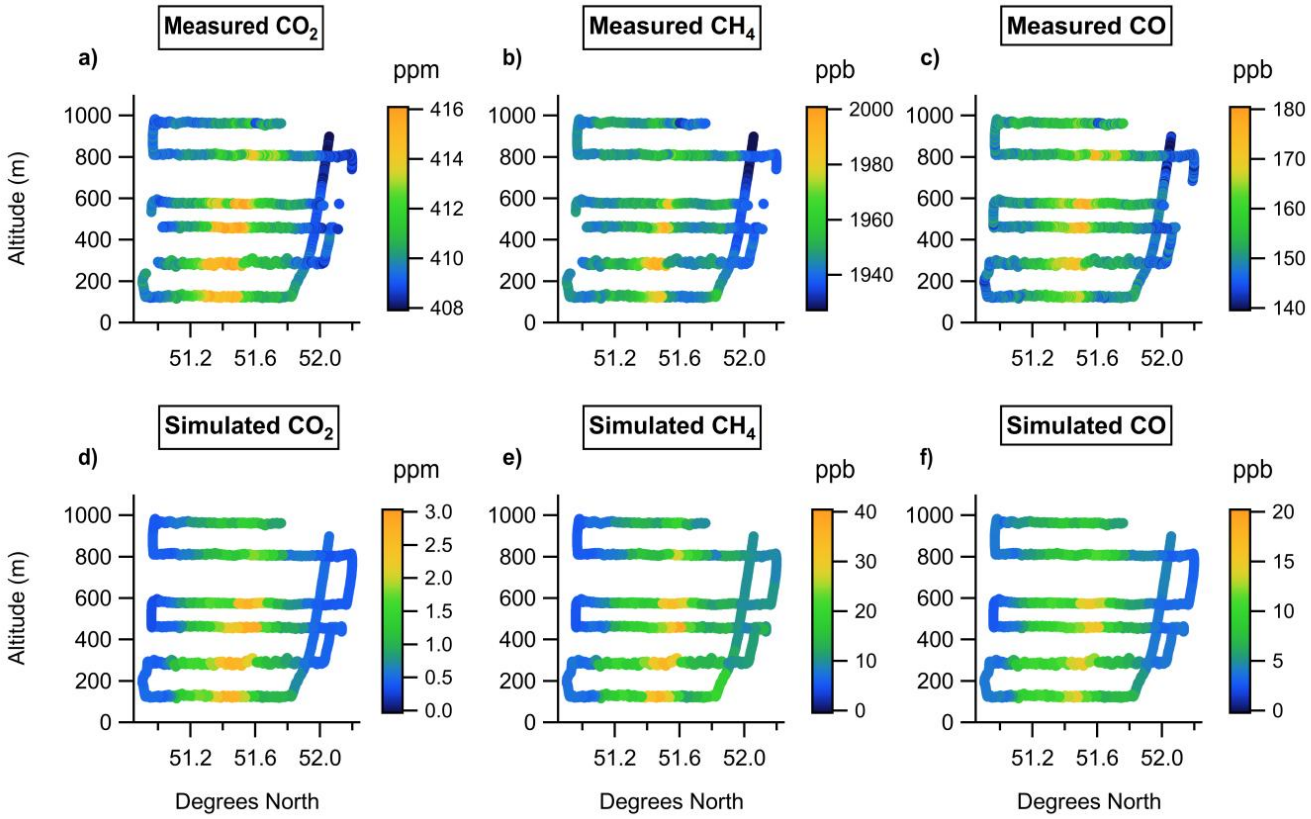

Figure 2: Altitude-latitude projections of measured mole fraction (a – c) and simulated (d – f) mole fraction enhancement downwind of London for each species.

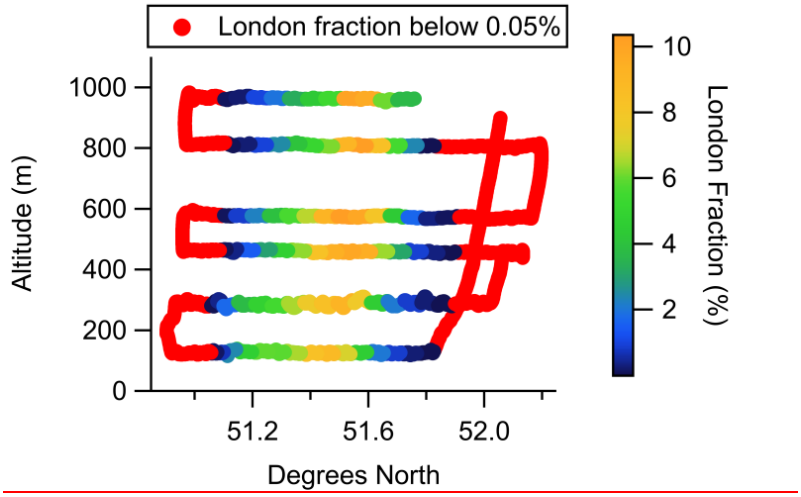

Figure 3: Altitude-latitude projection showing the influence of London on the downwind sampling, as determined from the NAME air histories. The colour scale represents the fraction of aggregated NAME air history D_{ij} within the Greater London administrative region for each NAME release period. Background periods, where the London fraction is less than 0.05%, are shown in red.

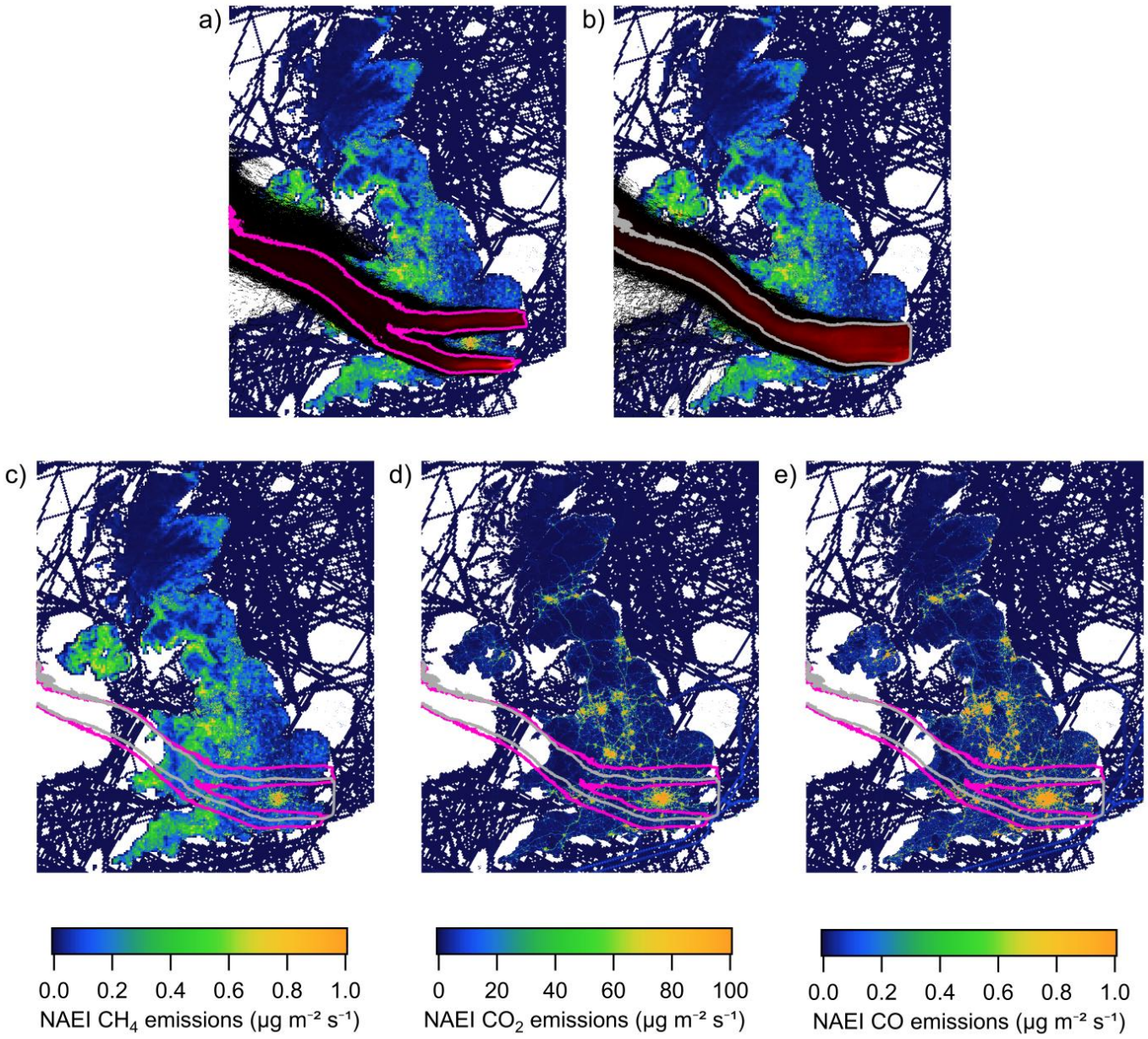

Figure 4: NAME air histories aggregated over (a) background sampling periods, and b) in-plume sampling periods, overlaid on an NAEI emissions inventory map for CH_4 (shown using a saturated colour scale). Both air histories have been normalised such that they sum to 1, with grey and pink contours shown in each plot surrounding the vast majority (99.9995%) of sample influence. These contours are included in panel c), d) and e) to provide a better visual comparison between the two aggregate air histories in the context of the inventory emissions for CH_4, CO_2 and CO respectively.

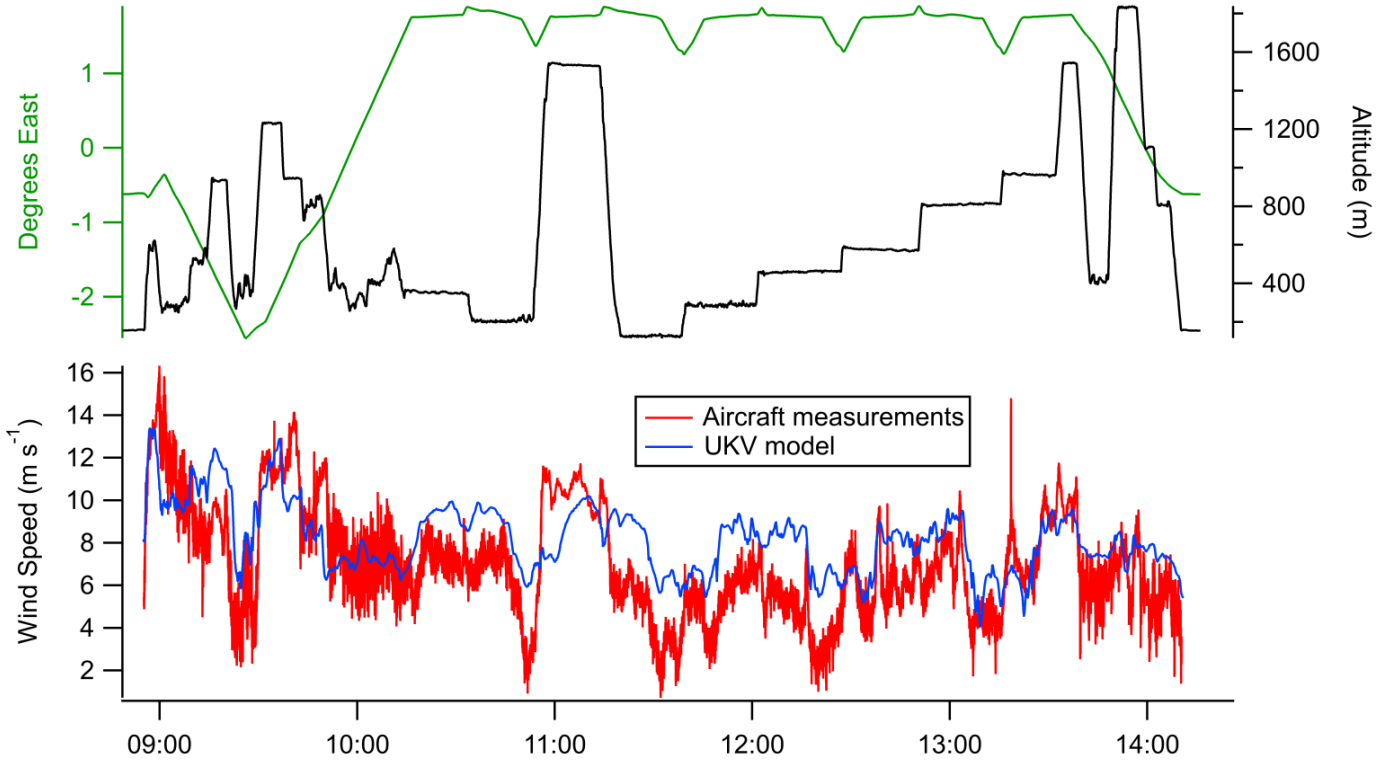

Figure 5: Comparison between wind speeds measured by the aircraft and the corresponding wind speeds at the aircraft location from the UKV model. It can be seen that the model generally overestimates wind speed within the boundary layer.

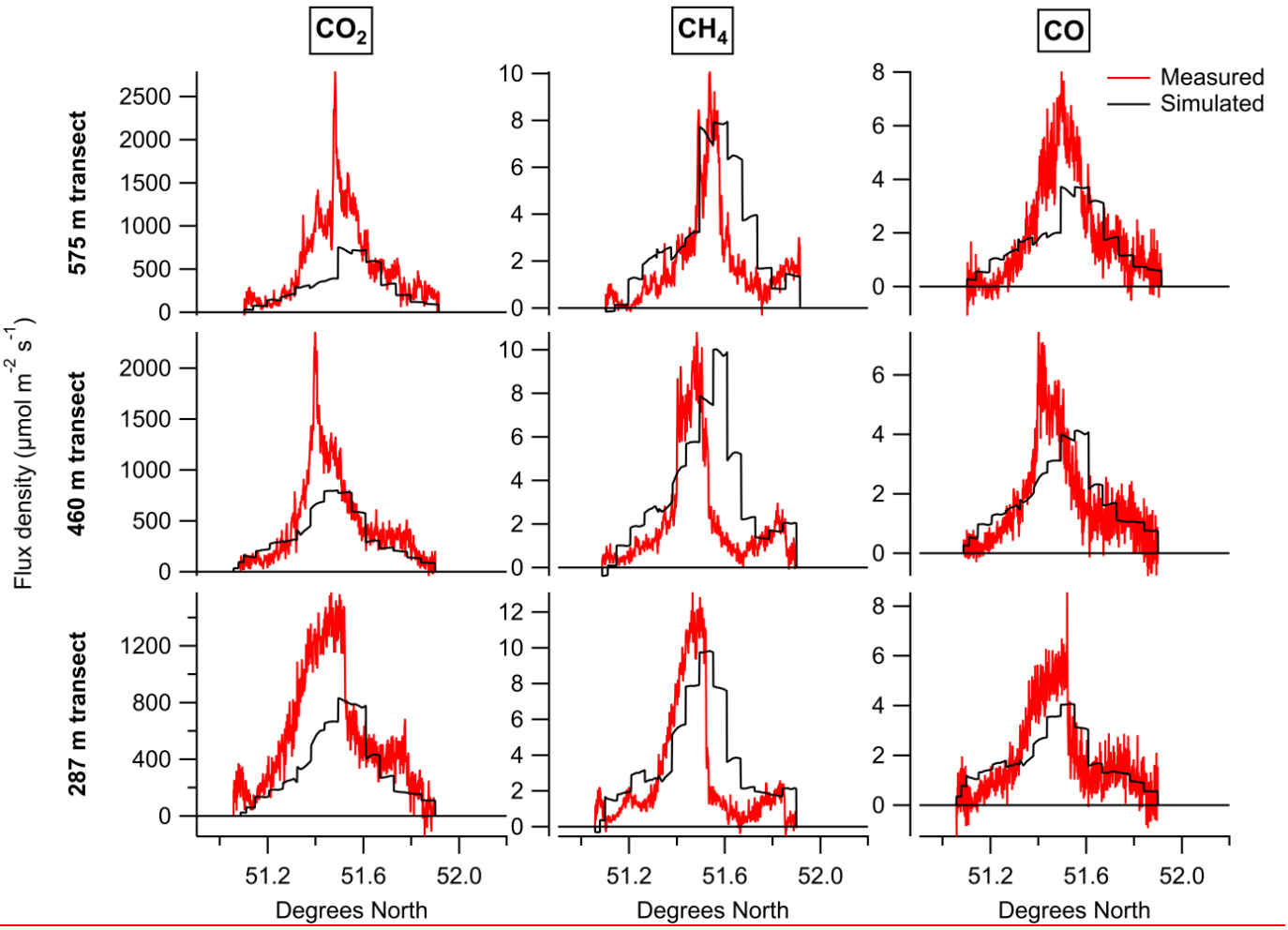

Figure 6: Measured and simulated flux densities for CO_2, CH_4 and CO, given for the three transects (287 m, 460 m and 575 m) used to assess inventory accuracy.

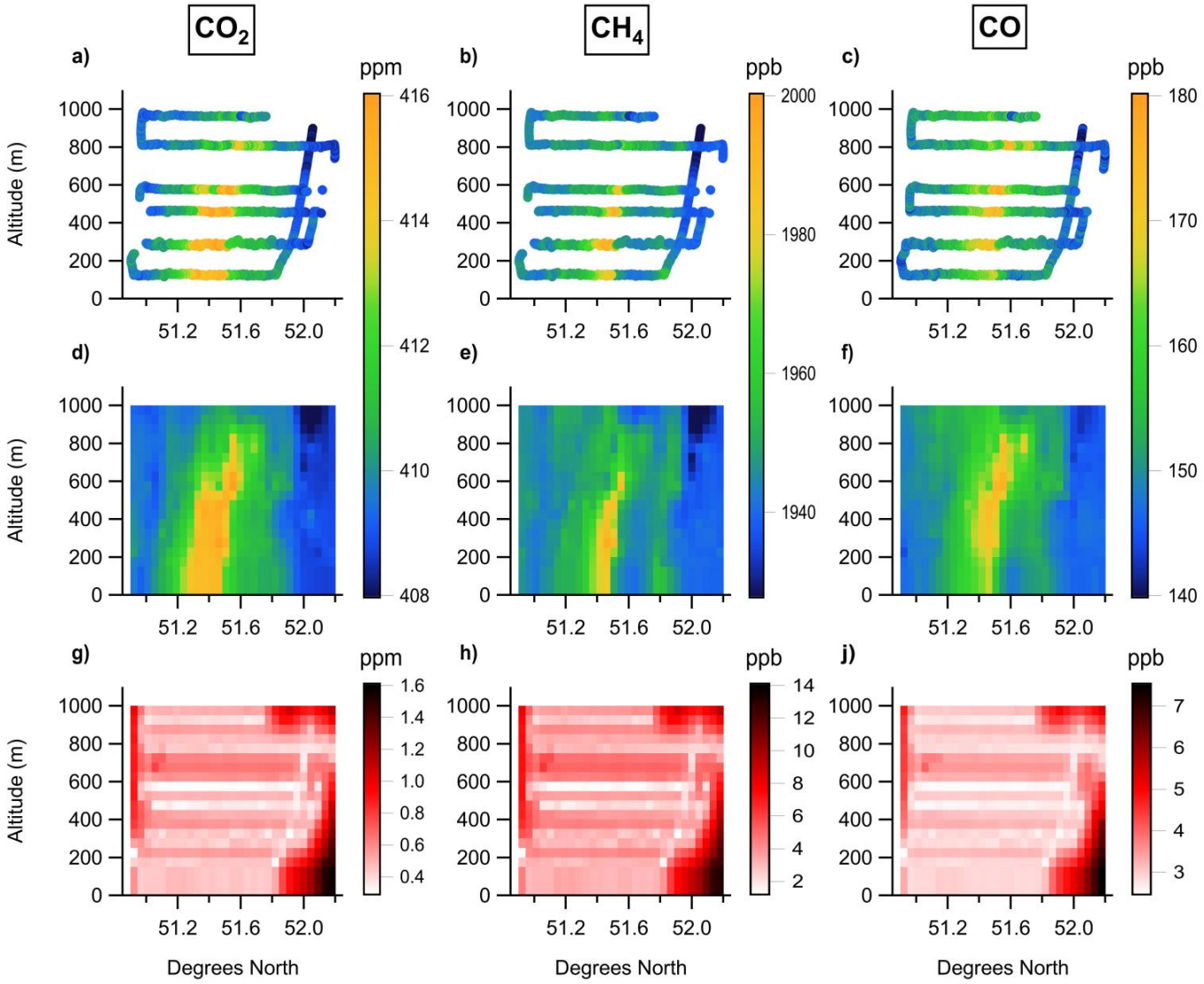

Figure 7: Altitude-latitude projections of: a) – c) measured data, d) – f) kriged data, g) – j) kriging standard error, for CO_2, CH_4 and CO respectively.

5